

What is claimed is:

1           1.       A method for predicting single nucleotide polymorphisms, comprising the  
2 steps of:

3           obtaining a variation predictiveness matrix; and

4           predicting one or more single nucleotide polymorphisms of a nucleic acid sequence  
5 based on the variation predictiveness matrix.

1           2.       The method of claim 1 further comprising one or more nucleic acid sequences  
2 with chemical modifications.

1           3.       The method of claim 2, wherein the chemical modifications include  
2 methylation or other chemical groups that incorporate additional charge, polarizability,  
3 hydrogen bonding, electrostatic interaction, and fluxionality to the individual nucleic acid  
4 bases or to the nucleic acid sequence as a whole.

1           4.       The method of claim 1, wherein the step of predicting the likelihood of one or  
2 more single nucleotide polymorphisms comprises the steps of:

3           comparing the nucleic acid sequence one or more bases at a time with the variation  
4 predictiveness matrix to assign a variation value to bases in the nucleic acid sequence; and

5           selecting the polymorphisms that will likely cause a variation in one or more bases of  
6 the nucleic sequence based on the variation value.

1           5.       The method of claim 4, wherein the variation in one or more bases is  
2 nonsynonymous.

1           6.       The method of claim 4, wherein the variation in one or more bases is  
2 synonymous.

1           7.       The method of claim 1, further comprising the step of generating a dataset of  
2 single nucleotide polymorphisms for one or more nucleic acid sequences.

1           8.       The method of claim 1, wherein the step of obtaining a variation  
2       predictiveness matrix, further comprises the steps of:

3           calculating a variation frequency from a first base to a second base in a dataset of two  
4       or more genes; and

5           generating the variation predictiveness matrix from the calculated variation  
6       frequency.

1           9.       The method of claim 8 wherein the dataset comprises genes with nucleic acid  
2       chemical modifications.

1           10.      The method of claim 9, wherein the chemical modifications include  
2       methylation or other chemical groups that incorporate additional charge, polarizability,  
3       hydrogen bonding, electrostatic interaction, and fluxionality to the individual nucleic acid  
4       bases or to the nucleic acid as a whole.

1           11.      The method of claim 8, wherein the variation frequency is determined from a  
2       known mutation dataset.

1           12.      The method of claim 8, wherein the variation frequency is determined from a  
2       dataset of known diseases.

1           13.      The method of claim 8, wherein the variation frequency is determined from a  
2       dbSNP database.

1           14.      The method of claim 8, wherein the variation frequency is determined from a  
2       non-human mutation database.

1           15.      The method of claim 8, wherein the variation frequency is determined from a  
2       disease-specific database.

1           16.      The method of claim 8, wherein the variation frequency is determined from a  
2       non-human disease database.

T0021-4058550

1 17. The method of claim 8, wherein the variation frequency is determined from a  
2 HGMD database.

1 18. The method of claim 8, wherein the variation frequency is determined from a  
2 linkage database.

1 19. The method of claim 8, wherein the variation frequency is determined from a  
2 splice variant database.

1 20. The method of claim 8, wherein the variation frequency is determined from a  
2 translocation database.

1 21. The method of claim 8, wherein the variation frequency is determined from a  
2 database of known mutations.

1 22. The method of claim 8, wherein the variation frequency is further adjusted for  
2 wild type genes.

1 23. The method of claim 8, wherein the variation frequency is further adjusted for  
2 engineered or non-naturally occurring genes.

1 24. The method of claim 8, wherein the variation frequency is further adjusted for  
2 conservative polymorphisms.

1 25. The method of claim 8, wherein the variation frequency is further adjusted for  
2 non-conservative polymorphisms.

1 26. The method of claim 8, wherein the variation frequency is further adjusted for  
2 cDNA stability.

1 27. The method of claim 8, wherein the variation frequency is further adjusted for  
2 predicted DNA structure.

1 28. The method of claim 8, wherein the variation frequency is further adjusted for  
2 predicted RNA structure.

1           29.     The method of claim 8, wherein the variation frequency is further adjusted for  
2     predicted protein structure.

1           30.     The method of claim 8, wherein the variation frequency is further adjusted for  
2     post-translational modification sequences.

1           31.     The method of claim 8, wherein the variation frequency is further adjusted for  
2     protein stability.

1           32.     The method of claim 8, wherein the variation frequency is further adjusted for  
2     predicted protein transport.

1           33.     The method of claim 8, wherein the variation frequency is further adjusted for  
2     shuffled genes.

1           34.     The method of claim 8, wherein the variation frequency is further adjusted for  
2     site-directed mutagenesis genes.

1           35.     The method of claim 8, wherein the variation frequency is further adjusted for  
2     methylated sequences

1           36.     The method of claim 8, wherein the variation frequency is further adjusted for  
2     epigenetic variation.

1           37.     The method of claim 8, wherein the nucleic acid sequence comprises a cDNA  
2     sequence.

1           38.     The method of claim 8, wherein the nucleic acid sequence comprises genomic  
2     sequence.

1           39.     The method of claim 8, wherein the nucleic acid sequence comprises an  
2     intron/exon boundary.

1           40.     The method of claim 8, wherein the nucleic acid sequence comprises a  
2     transcriptional control sequence.

TOOET-4066660

50. The method of claim 1, where the nucleic acid sequence comprises a human genome.

1           51.     The method of claim 1, where the nucleic acid sequence comprises a gene  
2 cluster for a target human disease.

1           52.     The method of claim 1, where the variation predictiveness matrix is based on  
2 a mutant gene dataset that comprises a human mutation database.

1           53.     The method of claim 1, wherein the steps are affected by a computer program.

1           54.     The method of claim 53, wherein the computer program is SNIDE.

1           55.     The method of claim 53, wherein the computer program is SNooP.

1           56.     The method of claim 1, wherein the variation predictiveness matrix is  
2 determined in silico from a human mutant database.

1           57.     The method of claim 1, wherein the step of predicting a likelihood of one or  
2 more single nucleotide polymorphisms is determined in silico.

1           58.     A method for creating a variation predictiveness value for use in a variation  
2 predictiveness matrix, comprising the steps of:

3                 calculating the variation frequency from a first nucleic acid to a second nucleic acid  
4 in a dataset of two or more variations; and

5                 determining a variation predictiveness value from the calculated variation frequency.

1           59.     The method of claim 58, further comprising the step of generating a variation  
2 predictiveness matrix that correlates the frequency of a first to a second variation with the  
3 variation predictiveness value.

1           60.     The method of claim 58, wherein the dataset comprises genes with nucleic  
2 acid chemical modifications.

1           61.     The method of claim 60, wherein the chemical modifications include  
2 methylation or other chemical groups that incorporate additional charge, polarizability,

3 hydrogen bonding, electrostatic interaction, and fluxionality to the individual nucleic acid  
4 bases or to the nucleic acid as a whole.

1 62. The method of claim 58, wherein the variation frequency is determined from a  
2 known mutation dataset.

1 63. The method of claim 58, wherein the variation frequency is determined from a  
2 dataset of known diseases.

1 64. The method of claim 58, wherein the variation frequency is determined from a  
2 dbSNP database.

1 65. The method of claim 58, wherein the variation frequency is determined from a  
2 non-human mutation database.

1 66. The method of claim 58, wherein the variation frequency is determined from a  
2 disease-specific database.

1 67. The method of claim 58, wherein the variation frequency is determined from a  
2 non-human disease database.

1 68. The method of claim 58, wherein the variation frequency is determined from a  
2 HGMD database.

1 69. The method of claim 58, wherein the variation frequency is determined from a  
2 linkage database.

1 70. The method of claim 58, wherein the variation frequency is determined from a  
2 splice variant database.

1 71. The method of claim 58, wherein the variation frequency is determined from a  
2 translocation database.

1 72. The method of claim 58, wherein the variation frequency is determined from a  
2 database of known mutations.

PATENT 40686650

1           73.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for wild type genes.

1           74.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for engineered or non-naturally occurring genes.

1           75.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for conservative polymorphisms.

1           76.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for non-conservative polymorphisms.

1           77.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for cDNA stability.

1           78.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for predicted DNA structure.

1           79.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for predicted RNA structure.

1           80.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for predicted protein structure.

1           81.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for post-translational modification sequences.

1           82.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for protein stability.

1           83.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for predicted protein transport.

1           84.     The method of claim 58, wherein the variation frequency is further adjusted  
2     for shuffled genes.



1           99.    The method of claim 58, wherein variations comprise a translational control  
2   sequence.

1 100. The method of claim 58, wherein variations comprise a transcriptional control  
2 sequence.

1 101. The method of claim 58, wherein variations comprise a splicing control  
2 sequence.

1 102. The method of claim 59, wherein in the variation predictiveness matrix is  
2 normalized for the nucleotide usage of a target organism.

1 103. The method of claim 59, wherein the variation predictiveness matrix is  
2 generated from a mutant gene dataset that comprises all mutant genes in a mutant gene  
3 database.

1 104. The method of claim 58, wherein the variation predictiveness matrix is  
2 generated from a mutant gene dataset that comprises all mutant genes in a mutant gene  
3 database minus the known mutant genes of the mutant gene dataset.

1 105. The method of claim 58, where the nucleic acid comprises one or more bases.

1 106. The method of claim 58, where the nucleic acid comprises DNA.

1 107. The method of claim 58, where the nucleic acid comprises RNA.

1 108. The method of claim 58, where the nucleic acid comprises a triplet.

1 109. The method of claim 58, The method of claim 16, where the nucleic acid  
2 comprises a codon.

1 110. The method of claim 58, The method of claim 16, where the nucleic acid  
2 comprises one or more non-sequence base modifications.

1 111. The method of claim 58, where the nucleic acid comprises modified nucleic  
2 acids.

1 112. The method of claim 58, wherein modified nucleic acids include methylation  
2 or other chemical groups that incorporate additional charge, polarizability, hydrogen

3 bonding, electrostatic interaction, and fluxionality to the individual nucleic acid bases or to  
4 the nucleic acid as a whole.

1 113. The method of claim 58, where the nucleic acid comprises an entire genome.

1 114. The method of claim 58, where the nucleic acid comprises a human genome.

1 115. The method of claim 58, where the nucleic acid comprises a gene cluster for a  
2 target human disease.

1 116. The method of claim 58, where the variation predictiveness matrix is based on  
2 a mutant gene dataset that comprises a human mutation database.

1 117. The method of claim 58, wherein the steps are affected by a computer  
2 program.

1 118. The method of claim 58, wherein the computer program is SNIDE.

1 119. The method of claim 58, wherein the computer program is SNooP.

1 120. The method of claim 58, wherein the variation predictiveness value is  
2 determined in silico from a human mutant database.

1 121. The method of claim 58, wherein the step of predicting a likelihood of one or  
2 more single nucleotide variation is determined in silico.

1 122. A method for creating a polymorphism predictiveness value for use in a  
2 mutation predictiveness matrix, comprising the steps of:

3 calculating the mutation frequency from a first codon to a second codon in a dataset  
4 of two or more mutant genes; and

5 determining a polymorphism predictiveness value from the calculated mutation  
6 frequency.

1 123. The method of claim 122, further comprising the step of generating a codon  
2 polymorphism predictiveness matrix that correlates the frequency of a first to a second codon  
3 mutation with the polymorphism predictiveness value.

1 124 The method of claim 122, wherein the dataset comprises nucleic acids with  
2 chemical modifications.

1 125 The method of claim 124, wherein the chemical modifications include  
2 methylation or other chemical groups that incorporate additional charge, polarizability,  
3 hydrogen bonding, electrostatic interaction, and fluxionality to the individual nucleic acid  
4 bases or to the nucleic acid as a whole.

1 126 The method of claim 122, wherein the mutation frequency is determined from  
2 a known mutation dataset.

1 127 The method of claim 122, wherein the mutation frequency is determined from  
2 a dataset of known diseases.

1 128 The method of claim 122, wherein the mutation frequency is determined from  
2 a dbSNP database.

1 129 The method of claim 122, wherein the mutation frequency is determined from  
2 a non-human mutation database.

1 130 The method of claim 122, wherein the mutation frequency is determined from  
2 a disease-specific database.

1 131 The method of claim 122, wherein the mutation frequency is determined from  
2 a non-human disease database.

1 132. The method of claim 122, wherein the mutation frequency is determined from  
2 a HGMD database.

1 133. The method of claim 122, wherein the mutation frequency is determined from  
2 a linkage database.

1           145.    The method of claim 122, wherein the mutation frequency is further adjusted  
2   for post-translational modification sequences.

1 146. The method of claim 122, wherein the mutation frequency is further adjusted  
2 for protein stability.

1 147. The method of claim 122, wherein the mutation frequency is further adjusted  
2 for predicted protein transport.

1 148. The method of claim 122, wherein the mutation frequency is further adjusted  
2 for shuffled genes.

1 149. The method of claim 122, wherein the mutation frequency is further adjusted  
2 for site-directed mutagenesis genes.

1 150. The method of claim 122, wherein the mutation frequency is further adjusted  
2 for methylated sequences

1 151. The method of claim 122, wherein the mutation frequency is further adjusted  
2 for epigenetic variation.

1 152. The method of claim 122, wherein the mutant genes comprise a cDNA  
2 sequence.

1 153. The method of claim 122, wherein the mutant genes comprise genomic  
2 sequence.

1 154. The method of claim 122, wherein mutant genes comprise an intron/exon  
2 boundary.

1 155. The method of claim 122, wherein mutant genes comprise exons.

1 156. The method of claim 122, wherein mutant genes comprise other SNPs.

1 157. The method of claim 122, wherein mutant genes comprise inversions.

1 158. The method of claim 122, wherein mutant genes comprise deletions.

1 159. The method of claim 122, wherein mutant genes comprise splice variations.



1 174. The method of claim 122, where the codon comprises one or more non-  
2 sequence base modifications.

1 175. The method of claim 122, wherein the codon further comprises modifications.

1 176. The method of claim 122, wherein modifications include methylation or other  
2 chemical groups that incorporate additional charge, polarizability, hydrogen bonding,  
3 electrostatic interaction, and fluxionality to the individual nucleic acid bases or to the nucleic  
4 acid as a whole.

1 177. The method of claim 122, where the codon comprises an entire genome.

1 178. The method of claim 122, where the codon comprises a human genome.

1 179. The method of claim 122, where the codon comprises a gene cluster for a  
2 target human disease.

1 180. The method of claim 122, where the codon polymorphism predictiveness  
2 matrix is based on a mutant gene dataset that comprises a human mutation database.

1 181. The method of claim 122, wherein the step of predicting a likelihood of one or  
2 more single nucleotide polymorphisms is determined in silico.

1 182. A method for creating a variation predictiveness matrix, comprising the steps  
2 of:

3 calculating the variation frequency from a first nucleic acid to a second nucleic acid  
4 in a dataset of two or more variations;

5 determining a variation predictiveness value from the calculated variation frequency;  
6 and

7 generating a variation predictiveness matrix that correlates the frequency of a first to  
8 a second nucleic acid with the variation predictiveness value.

DOCKET 4068660



1 183. The method of claim 182, wherein the dataset comprises nucleic acids with  
2 chemical modifications.

1 184. The method of claim 183, wherein the chemical modifications include  
2 methylation or other chemical groups that incorporate additional charge, polarizability,  
3 hydrogen bonding, electrostatic interaction, and fluxionality to the individual nucleic acid  
4 bases or to the nucleic acid as a whole.

1 185. The method of claim 182, wherein the variation frequency is determined from  
2 a variation dataset.

1 186. A method for creating a polymorphism predictiveness matrix, comprising the  
2 steps of:

3 calculating the mutation frequency from a first codon to a second codon in a dataset  
4 of two or more mutant genes;

5 determining a polymorphism predictiveness value from the calculated mutation  
6 frequency; and

7 generating a codon polymorphism predictiveness matrix that correlates the frequency  
8 of a first to a second codon mutation with the polymorphism predictiveness value.

1 187. The method of claim 186, wherein the dataset comprises nucleic acids with  
2 chemical modifications.

1 188. The method of claim 187, wherein the chemical modifications include  
2 methylation or other chemical groups that incorporate additional charge, polarizability,  
3 hydrogen bonding, electrostatic interaction, and fluxionality to the individual nucleic acid  
4 bases or to the nucleic acid as a whole.

1 189. The method of claim 186, wherein in the codon polymorphism predictiveness  
2 matrix is normalized for the codon usage of a target organism.

DOCKET # 40685550

1 190. The method of claim 186, wherein the codon polymorphism predictiveness  
2 matrix is generated from a mutant gene dataset that comprises all mutant genes in a mutant  
3 gene database.

1 191. The method of claim 186, wherein the codon polymorphism predictiveness  
2 matrix is generated from a mutant gene dataset that comprises all mutant genes in a mutant  
3 gene database minus the known mutant genes of the mutant gene dataset.

1 192. The method of claim 186, wherein the codon comprises one or more bases.

1 193. The method of claim 186, where the codon comprises a triplet.

1 194. The method of claim 186, where the codon comprises a codon.

1 195. The method of claim 186, where the codon comprises one or more non-  
2 sequence base modifications.

1 196. An isolated and purified nucleic acid comprising a predicted single nucleotide  
2 variation of a nucleic acid sequence based on the variation predictiveness matrix sequence of  
3 claim 1.

1 197. An isolated and purified nucleic acid comprising a predicted single nucleotide  
2 polymorphism of a wild-type gene sequence based on the codon mutation predictiveness  
3 matrix sequence of claim 1.

1 198. An apparatus for detecting a single nucleotide polymorphism comprising:  
2 a substrate; and

3 one or more isolated and purified nucleic acids comprising a predicted single  
4 nucleotide variation of a nucleic acid sequence based on a variation predictiveness matrix  
5 sequence affixed to the substrate.

1 199. The apparatus of claim 198, wherein the substrate comprises a  
2 microfabricated solid surface to which molecules may be attached through either covalent or  
3 non-covalent bonds.

1           200. The apparatus of claim 198, wherein the substrate further comprises  
2           Langmuir-Bodgett films, glass, functionalized glass, germanium, silicon, PTFE, polystyrene,  
3           gallium arsenide, gold, silver, or any materials comprising amino, carboxyl, thiol or hydroxyl  
4           functional groups incorporated on a planar or spherical surface.

1           201. An apparatus for detecting a single nucleotide polymorphism comprising:  
2           a substrate; and  
3           one or more isolated and purified nucleic acids comprising a predicted single  
4           nucleotide polymorphism of a wild-type gene sequence based on a codon polymorphism  
5           predictiveness matrix. sequence affixed to the substrate.

1           202. The apparatus of claim 201, wherein the substrate comprises a  
2           microfabricated solid surface to which molecules may be attached through either covalent or  
3           non-covalent bonds.

1           203. A computer program embodied on a computer readable medium for predicting  
2           variations, comprising:

3           a code segment for creating variation predictiveness matrix from a nucleic acid  
4           dataset;

5           a code segment for comparing a wild-type gene sequence with the variation  
6           predictiveness matrix; and

7           a code segment for predicting variations in the wild-type gene sequence based on the  
8           comparison.

1           204. A computer program embodied on a computer readable medium for predicting  
2           polymorphisms, comprising:

3           a code segment for creating a codon mutation predictiveness matrix from a mutant  
4           gene dataset;

5 a code segment for comparing a wild-type gene sequence with the codon  
6 polymorphism predictiveness matrix; and

7 a code segment for predicting polymorphisms in the wild-type gene sequence based  
8 on the comparison.

1 205. A polymorphism prediction dataset, comprising:

2 a first nucleic acid;

3 a second nucleic acid variation that correlates to a polymorphism from the first  
4 nucleic acid; and

5 a variation predictiveness value determined from known variations in a variation  
6 database for a target organism.

1 206. A polymorphism prediction dataset, comprising:

2 a first codon;

3 a second codon mutation that correlates to a mutation from the first codon; and

4 a codon polymorphism predictiveness value determined from known mutations in a  
5 mutation database for a target organism.

1 207. A single nucleotide polymorphism determined by the method of claim 1.

1 208. A method for predicting single nucleotide polymorphisms, comprising the  
2 steps of:

3 inputting each codon in a queried nucleic acid sequence;

4 determining each possible nonsynonymous mutation;

5 assigning a predictiveness value to that mutation based on the identity of the wild-  
6 type and resultant codon; and

1           213.   An isolated and purified nucleic acid of claim 211, wherein the SNP is Thr-  
2   >Met substitution in BDKRB2 at position 383.